Analyzing Relationships Among Features of Diabetes-Induced Cerebral Microvascular Disease Using Causal Inference Methods

Lauren Shen

BMED027

Abstract

Type 2 diabetes mellitus (T2DM) is a debilitating condition affecting 462 million people worldwide and puts those affected at risk of developing more severe diseases, including Alzheimer's and stroke. T2DM is strongly linked to cerebral microvascular disease (CVD), which drastically decreases the brain's blood flow and oxygen supply, increasing the risk of the aforementioned diseases. However, while features of CVD have been previously investigated, the exact relationships among these features are unknown, and how one feature may cause another is uncertain. Defining more precise connections among these features would allow physicians to better detect CVD-induced diseases and enable earlier treatment for patients. In this work, causal inference methods of machine learning were applied to better understand relationships between diabetes severity, gender, inflammation, and cerebral hypoperfusion. Specifically, a Bayesian network algorithm and a new method utilizing logic from Granger causality were implemented on data from subsets of a Physionet data set. Diabetes severity and specific inflammation biomarkers were found to be significant, and differences in cerebral vasoreactivity correlations were shown in males and females. While causal inference methods have not been extensively applied to this particular issue previously, this work shows that these methods have extreme potential for uncovering relationships in features of CVD, aiding medical practitioners in their treatments for those suffering from a T2DM- or CVD-induced disease.

1 Introduction

Type 2 diabetes mellitus (T2DM or DM) is a debilitating condition affecting 462 million people worldwide that can severely alter the brain, disrupting cognitive function and leading to other diseases [1]. In particular, T2DM can cause cerebral microvascular disease (CVD), also known as cerebral small vessel disease (SVD). CVD is associated with cerebral hypoperfusion, high blood pressure, hyperglycemia, inflammation, and obesity, and severely reduces blood flow and oxygen delivery in the brain. It has also been linked to a myriad of other conditions, including Alzheimer's, dementia, stroke, and cognitive decline. While CVD alone can result in the aforementioned conditions, T2DM-induced CVD increases the risk of these illnesses and has more severe effects for patients when compared to those affected solely by CVD.

While the features that are related to CVD have been previously investigated, the exact relationship among these features remains unclear. Possible risk factors of CVD include T2DM, cardiovascular disease, and smoking [2]. Also, CVD has been linked to significant changes in the brain, such as inflammation and hypoperfusion (insufficient oxygen delivery). Additionally, even though gender has not been seriously considered as a possible risk factor for CVD, it is known that stroke, a likely effect of CVD, has different effects on men and women [3]. If the relationship between these causes and features of CVD becomes more clear, it is likely that patients suffering from T2DM or related conditions will be able to receive better treatment and testing for issues that may not be immediately apparent, such as cognitive decline. Figure 1 shows the features of CVD, including how causes, symptoms, and affects are related.

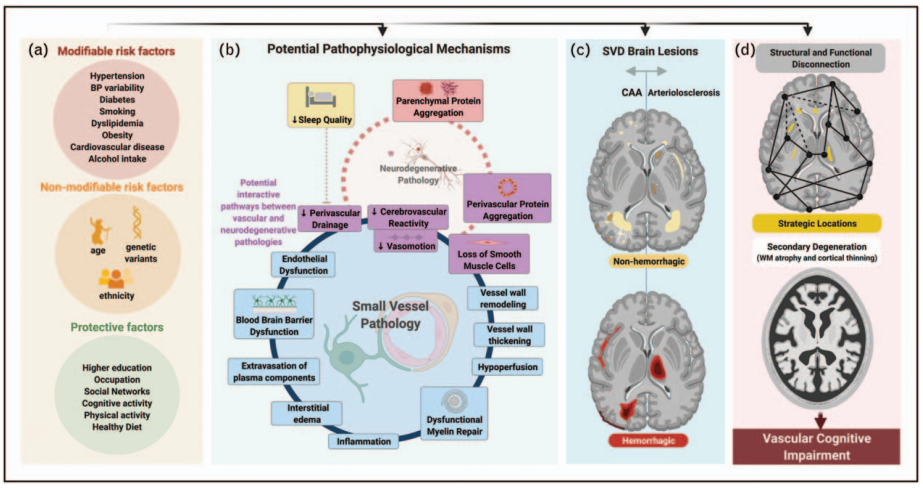

Figure 1: The diagram shows the causes, symptoms, and effects of CVD [2]. Inflammation and hypoperfusion are included in the middle-left section. Note that diabetes is included as a cause on the left while gender is not.

However, analyzing these features from a biological perspective is difficult; for example, understanding pathways in the brain at the molecular level for inflammation biomarkers would be an involved process. Furthermore, symp-

toms of CVD other than inflammation and hypoperfusion could be taken into account, which would significantly complicate a biological approach. While it is difficult to directly study these features and their relationship in patients, taking a machine learning perspective to explore these features using data collected from these patients is a promising approach. To better understand the relationship among severity of T2DM, gender, inflammation, and cerebral hypoperfusion, we utilized a variety of statistical methods and machine learning tools to investigate possible causal relationships among the features and further analyze any connections found. Specifically, we implemented a Bayesian network model, linear regression, and a method inspired by Granger causality to explore the relationship between pairs of features and compared the three methods to find any commonalities.

While machine learning techniques have been previously applied to the medical field, such as in diagnosing or predicting a particular disease based on a patient's symptoms, these tools have not been extensively applied in this area. Specifically, a causal inference method examining the connection between T2DM and CVD has not been previously implemented. We utilize both Bayesian networks and the novel Granger-inspired method to approach any connections among features from different angles and highlight any notable relationships. Additionally, the Granger-inspired method borrows logic from Granger causality (which is traditionally applied to time series data) and seeks to find a causal relationship between data that is not dependent on time. The method utilizes other machine learning regression techniques such as linear regression or support vector machines (SVMs) to execute its logic.

Our research seeks to address the following questions:

1. How are severity of T2DM, gender, inflammation biomarkers, and cerebral hypoperfusion (measured using cerebral vasoreactivities) related in the context of CVD? Furthermore, how can machine learning algorithms be applied to this issue to further explore and define the relationship among these features? Applying these algorithms can allow us to discover new connections between pairs of features that may have been difficult to find and investigate using a strictly biological method.

2. How can further knowledge of the relationships between the features and causes of CVD aid patients of related diseases, such as T2DM, stroke, Alzheimer's, and dementia? If connections among features are known, it is highly likely that this information can be used to detect these diseases early and aid patients in treatment as well.

2 Literature Review

2.1 Previous Studies

Previous research has addressed many features of T2DM and CVD. Features of CVD can be considered causes, symptoms, or effects: causes include T2DM, gender, and BMI [4]; symptoms include inflammation, decreases in cerebral blood flow, and hypoperfusion; and effects include gait speed, cognitive decline, stroke, and Alzheimer's. Many studies relating to two or more of these features have been conducted. For example, studies have been conducted at the

intersection of T2DM and each of gait speed [5], inflammation [6], and low cerebral vasoreactivity [6]. Additionally, studies have also explored the relationship between cognitive decline and each of hypoperfusion [7], inflammation [6], and blood pressure [8]. A study also investigated the different effects of ischemic stroke in the brains of males and females [3]. However, the exact relationship among these features in the underlying context of CVD remains unknown, and few studies have applied machine learning methods of causal inference to analyze the features of the intersection of T2DM and CVD. Figure 2 shows how the causes, symptoms, and effects of T2DM are related.

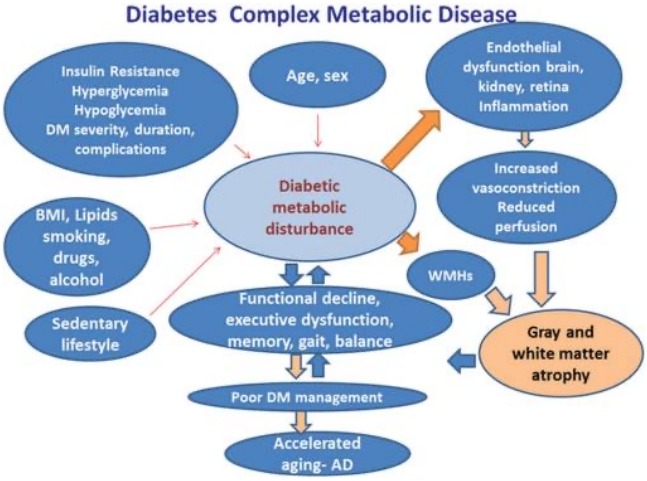

Figure 2: The flow chart shows the causes, symptoms, and effects of T2DM [9]. Note that many effects of T2DM are similar to symptoms or effects of CVD.

Our central research question involves the relationships among the following four features of cerebral microvascular disease (CVD): severity of T2DM, gender, inflammation biomarkers, and cerebral vasoreactivities. A distinction in the nature of these four factors arises; T2DM severity and gender can be considered causes of CVD, while inflammation and cerebral vasoreactivities are symptoms of CVD. While our central question revolves around these causes and symptoms, an effect of CVD, gait speed, is also a notable feature to be considered.

2.2 Features of CVD

2.2.1 Severity of T2DM

Severity of T2DM is measured using a biomarker known as HbA1c, or hemoglobin A1c. An HbA1c test measures the amount of glucose currently attached to the blood's hemoglobin, and the percentage of HbA1c can be used to suggest if a patient is prediabetic or diabetic [10]. Table 1 shows the levels of HbA1c % that correspond to prediabetes or diabetes.

2.2.2 Gender

The significance of gender in this study is related to the different effects of ischemic stroke in the brains of males and females [2]. CVD is a cause of stroke

Table 1: The percentages of HbA1c % used to determine if a patient is not diabetic, is prediabetic, or is diabetic. **Unless otherwise noted, all graphs and charts were created by the student researcher.**

Normal	Prediabetic	Diabetic
HbA1c % $<$ 5.7%	5.7 % $<$ HbA1c % $<$ 6.4 %	6.4 % $<$ HbA1c %

and affects features of the brain that have a role in stroke, such as blood flow and blood pressure. Thus, since gender has been shown to be a factor in the effects of stroke, it is possible that CVD may have different effects for males and females as well.

2.2.3 Inflammation Biomarkers

The inflammation biomarkers analyzed included SVCAM, SICAM, CRP, IL-6, and TNF alpha.

1. SVCAM (Soluble Vascular Cell Adhesion Molecule-1, or sVCAM-1) is related to hypertension, cerebral inflammation, and endothelial cell dysfunction, which is another known feature of CVD. SVCAM was found to be related to slower walking speed and can also be used as an indicator of CVD [11].

2. SICAM (Soluble Intercellular Adhesion Molecule 1, or sICAM-1) is also related to endothelial cells. Additionally, SVCAM and SICAM have been studied together [12] and the two biomarkers have been found at higher levels in diabetic children, showing a significant connection to T2DM in general.

3. CRP (C-reactive protein) is a protein released in the body shortly after inflammation and is used as a non-specific inflammation biomarker. CRP tests are conducted to determine if a patient has inflammation and in some cases is used as an aid in detecting or diagnosing certain diseases [13].

4. IL-6 (Interleukin 6) can be a sign of chronic inflammation and plays a significant role in the immune system's response to inflammation [14].

5. TNF alpha (Tumour Necrosis Factor alpha) is produced during acute inflammation and is also important in cell signalling pathways. The biomarker is significant in its relation to endothelial cells, infection, and cancer [15].

2.2.4 Global and Cerebral Vasoreactivities

Vasoreactivity is defined as "the capacity to regulate cerebral blood flow in response to carbon dioxide challenges" [5]. Thus, cerebral vasoreactivity is strongly related to blood flow and hypoperfusion, making it an appealing feature of CVD to examine. Global vasoreactivity is a measure of the vasoreactivity

in the brain as a whole. In the studies we acquired our data sets from, cerebral vasoreactivities were measured for four lobes of the brain (frontal, parietal, temporal, and occipital) under three different breathing patterns (baseline, rebreathing, and hyperventilation), each of which involved a different amount of carbon dioxide. Figure 3 shows the differences in global vasoreactivities between the brain of a T2DM participant and that of a control participant as they exhibit different breathing patterns.

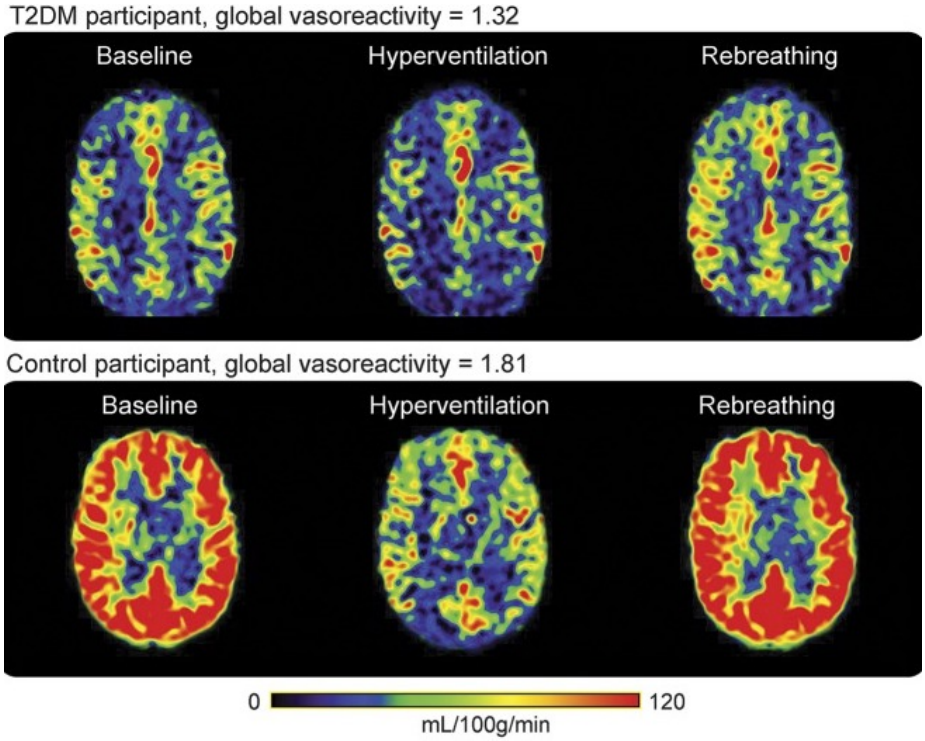

Figure 3: Differences in global vasoreactivities arise in the MRI scans between the brain of a T2DM participant and the brain of a control participant as the participants exhibit different breathing patterns, each pattern involving a different amount of carbon dioxide [6]. The T2DM participant has a much lower overall global vasoreactivity.

2.2.5 Walking Speed

Walking speed, also known as gait speed, has been shown to be related to cerebral vasoreactivity [5]. Our research shows walking speed is related to other features as well. Walking speed is generally slower in patients with T2DM than those without, likely due to the fact that T2DM affects cerebral vasoreactivity as well as other features in the brain, such as blood flow velocity. In our study, dual walking speed is also analyzed, which is the walking speed measured when the participant is focused on a mentally-stimulating task rather than solely on walking.

2.3 Machine Learning Techniques

Machine learning tools have been previously applied to studies related to our research. For example, a machine learning pipeline method has been implemented to study the association of cerebral perfusion with T2DM [16]. Machine learning has also been utilized to create a predictive model for T2DM [17].

2.3.1 Bayesian Networks

Bayesian networks are a method of causal inference, which can be described as a field of machine learning that seeks to determine causal relationships among features. While there are different algorithms of Bayesian networks, including the exact, exact A*, and greedy algorithms, we have used the Chow-Liu algorithm, which seeks to find a structured tree-like graph with the maximum probability from the data [18].

In a network or graph generated by a Bayesian network algorithm, features or variables are represented by *nodes*, and *edges* connecting nodes stem from the cause and point to the effect, effectively representing the causal relationship. Finally, each node contains a *conditional probability distribution*, which the model uses to construct the graph [19]. Figure 4 shows an example of a Bayesian network graph.

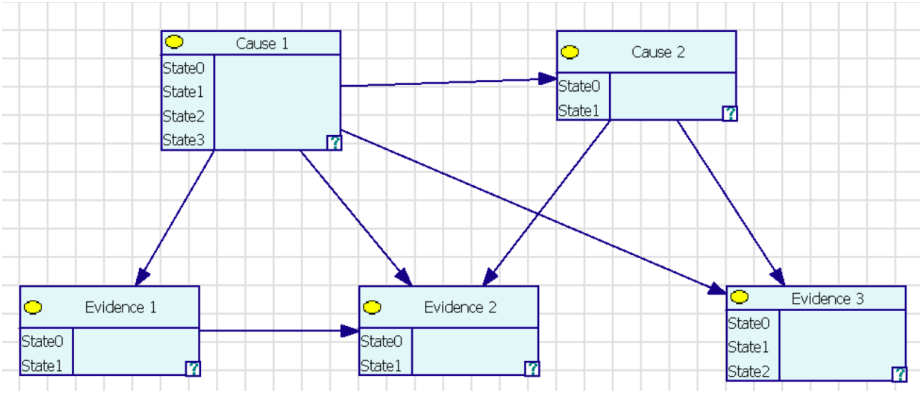

Figure 4: Example Bayesian network graph [19].

2.3.2 Granger-Inspired Method

Granger causality is a method of causal inference applied to time series data [20]. The reasoning behind Granger causality is as follows: If a factor X is determined to "Granger-cause" a factor Y, future values of Y can be predicted with past information from X better than utilizing past information of Y alone. Similarly, our Granger-inspired method seeks to find a relationship between two factors, A and C, with an intermediate factor B. First, we train a machine learning regression model, such as a linear regression model or an SVM, to find the error when using B to predict A. Then, using the same machine learning tool, we train a model to find the error when using B and C to predict A. If the error has decreased, it strongly suggests that C may have an influence on A or cause A.

3 Methodology

3.1 Data

3.1.1 Data Acquisition

We acquired data from three related studies regarding T2DM and its connection to cerebral vasoregulation, cerebral perfusion and cognitive decline, and cerebromicrovasular disease in the elderly. The three data sets, respectively named GE 71 [21, 22], GE 75 [23, 24], and GE 79 [25, 26], were obtained from Physionet [27]. As each data set contained data for distinct features, we combined the data sets and applied machine learning algorithms to specific subsets. GE 79 contained the most information for each patient, and was thus the subset used in the majority of the Bayesian network algorithms. However, since the study for GE 79 conducted research on patients during a time span of two years, GE 79 also contained data for each patient's follow-up visit. Thus, we split GE 79 further into two subsets for use in our algorithms and will refer to the subset of patients for the first year as GE 79-1 and the subset of patients in the second year as GE 79-2. Tables 2 and 3 show detailed distributions of the different subsets of the data and GE 79-1, respectively.

Table 2: The number of patients of each type in each subset of the data. There are generally more females than males and more T2DM group members than control group members.

Subset	Male	Female	Control	T2DM	All
GE 79-1	23	49	29	43	72
GE 71, 75, 79-1	101	108	62	147	209
GE 75, 79-1	72	86	42	116	158

Table 3: The number of patients of each type in GE 79-1. While it is a smaller subset of the data, GE 79-1 contains measures of many useful features for each patient.

	Male	Female	All
Control	8	21	29
T2DM	15	28	43
All	23	49	72

3.1.2 Data Preprocessing

Three methods of cleaning missing data were tested in the data preprocessing step to determine which would be most effective. The third method was ultimately used as it was found to better preserve the distribution of the data and produce consistent results with the machine learning algorithms used.

1. Filling in missing values with the mean across each column in the data set was considered as a method of cleaning the data. However, it was found that this heavily skewed the distribution in many columns and also neglected the possible differences in the distributions for each gender and group (control group or T2DM group). Thus, we attempted to use the second method instead.

2. We used numpy's randn function to generate random values from the standard normal distribution by providing the function with the mean and standard deviation of the column. While this method preserved the distribution of a column, it was found to produce inconsistent results when used with machine learning algorithms. The third method was found to alleviate these issues.

3. Finally, we filled in missing values with each column's mean conditioned on the group and gender of the patient. After splitting the data into four groups based on whether the patient was male or female and whether the patient was in the control or T2DM group, we filled in missing values by taking the mean of each column in each group. After the four groups were cleaned and assembled into the larger data set, it was found that this preserved the distribution fairly well and also produced consistent results. Thus, this method was chosen to clean the data set.

3.2 Pearson's r and Correlation Heatmaps

After preprocessing the data, we used linear regression to determine the Pearson's r value between any two chosen features in order to gain a sense of which features showed a connection prior to running a Bayesian network algorithm. This was performed using the linregress function from the scipy.stats Python package on GE 79-1. This particular data set was chosen as it contained the most information per patient. We found slight variations when taking into account differences in group or gender. Noting these differences, we decided to visualize the data. Using the corr() function from pandas and the heatmap() function from seaborn, we plotted correlation heatmaps and confirmed several differences in correlation for cerebral vasoreactivities and inflammation biomarkers in males and females.

3.3 Bayesian Networks

3.3.1 pomegranate Python Package

All machine learning algorithms were used within the Google Colab coding environment. The Python package pomegranate (used to implement Bayesian network algorithms) was installed using pip. Additionally, visualizing graphs involved installing pygraphviz. Thus, pygraphviz was installed with pip, and a supporting package, libgraphviz-dev, was installed with apt.

3.3.2 Chow-Liu Bayesian Network Algorithm

Using pomegranate, we implemented a Chow-Liu Bayesian network model on various subsets of the data set. While the Chow-Liu algorithm is not the most

accurate Bayesian network algorithm, it is the fastest, which was a necessary factor for use in this particular coding environment.

The Chow-Liu algorithm was first implemented on GE 79-1. This subset contained data regarding each patient's gender and group, severity of T2DM (measured by HbA1c %), measures of various inflammation biomarkers, global vasoreactivity, walking and dual walking speed, and cerebral vasoreactivities for four lobes of the brain during three different breathing patterns. To more closely examine the relationship among different features, we then ran the algorithm on smaller subsets of GE 79-1. Afterwards, to determine if the relationships found in GE 79-1 could be generalized to the larger data set, we ran the algorithm on subsets of GE 71, 75, and 79-1. Table 4 shows the specific features each subset of the data contained.

Table 4: The specific features each subset of the data contained.

Subset	Features
GE 79-1	Group, Gender, Severity, Inflammation Biomarkers, Global Vasoreactivity, Walking and Dual Walking Speeds, Cerebral Vasoreactivities
GE 71, 75, 79-1	Group, Gender, Severity, Walking Speed
GE 71, 75, 79-1	Group, Gender, Severity, Walking Speed, Cerebral Vasoreactivities
GE 75, 79-1	Group, Gender, Severity, Walking Speed, Global Vasoreactivity
GE 75, 79-1	Group, Gender, Severity, Inflammation Biomarkers, Walking Speed, Global Vasoreactivity

3.4 Granger-Inspired Method

Using the Granger-inspired method, we explored the significance of the connections found using the Bayesian network algorithm. Since the Granger-inspired method takes three factors into consideration (A, B, and C) to determine a causal relationship, we iterated through permutations of three factors at a time using the permuations package from the itertools Python library. These factors only included continuous data as it would be difficult to apply the regression models to discrete data, such as group or gender. Thus, while we did not consider group or gender, we did consider the relationships among the severity of T2DM, inflammation biomarker levels, and walking speed.

We utilized the subset of the data GE 71, 75 and 79-1, which contained the largest amount of patients. First, the data was split into training and test data.

Then, after predicting A using B (the AB group) and predicting A using B and C (the ABC group), the ratio between the mean squared error (MSE) of the AB group to the MSE of the ABC group was calculated. A ratio greater than 1 suggests that C has a causal influence over A, as it indicates that the error induced by the model decreased for the ABC group as a result of including C. We deemed results significant if the factors A, B, and C led to higher values for this ratio, indicating a greater decrease in error. This process was implemented using a linear regression model and again using an SVM method.

4 Results and Discussion

4.1 Correlation Heatmaps

Figures 5 and 6 can be compared to see differences in correlation among cerebral vasoreactivities in the data subset (GE 79-1) as a whole, the control group, and the T2DM group. Additionally, males showed a much stronger correlation (around 2 times more) among measures of cerebral vasoreactivities and a slightly stronger correlation among inflammation biomarkers when compared to female patients and to the entire data subset (Figures 7, 8, 9). Differences in inflammation biomarkers are most clearly seen in males in the control group (Figure 8), and differences in cerebral vasoreactivity correlation become more pronounced when comparing males and females in the T2DM group (Figure 9).

A possible explanation for why these differences in the male and female brains occur is related to the differing effects of ischemic stroke on the two genders. Ischemic stroke is related to blood clots and thus blood flow and cerebral vasoreactivity, and it is known to produce different effects for males and females.

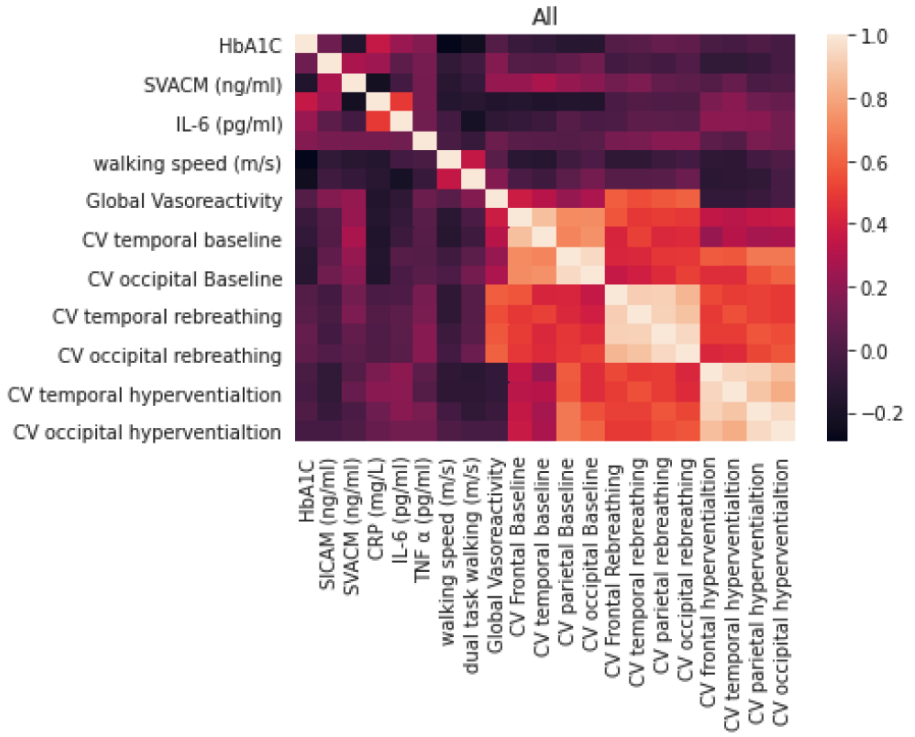

Figure 5: Correlation heatmap generated using data from all patients in GE 79-1. More correlation is shown among the cerebral vasoreactivities than other factors.

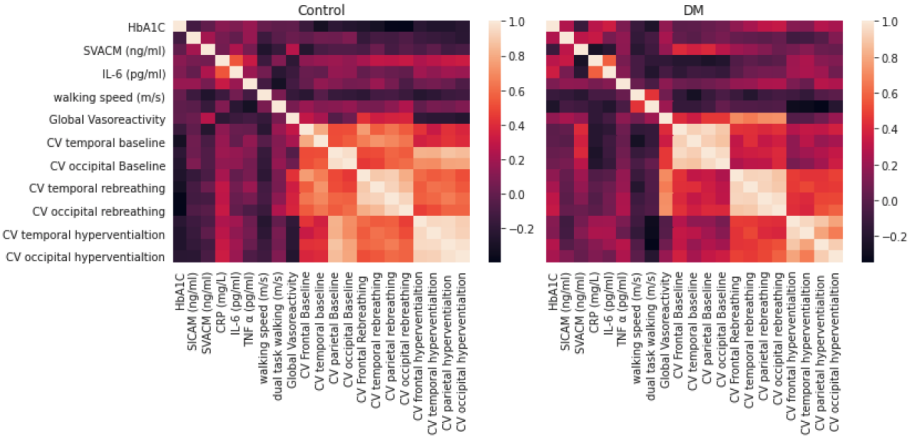

Figure 6: Comparing correlation heatmaps generated using data from control (left) and T2DM (right) patients in GE 79-1.

However, this is likely a result of sex-specific inflammatory signalling and immune responses, which were found to be an explanation for differences in ischemic stroke [3]. In particular, the SVCAM inflammation biomarker is known

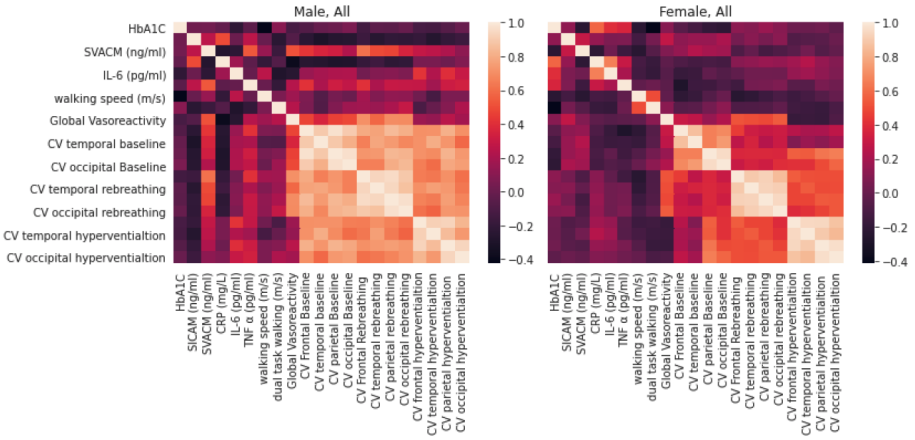

Figure 7: Comparing correlation heatmaps generated using data from male (left) and female (right) patients in GE 79-1.

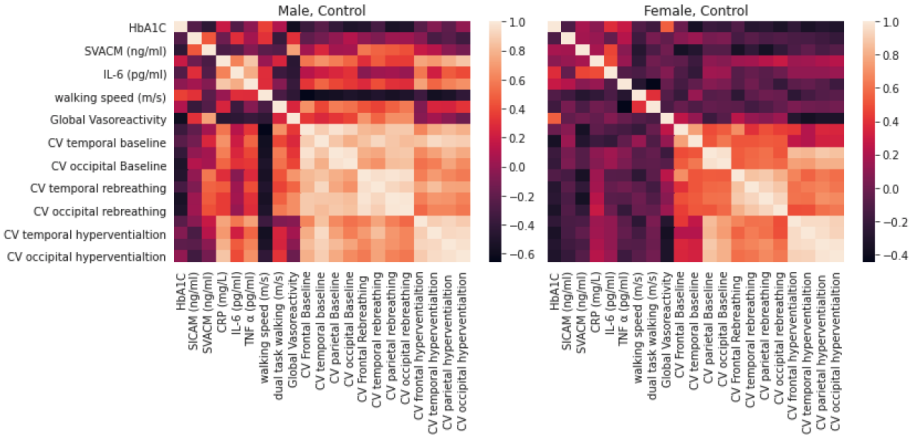

Figure 8: Comparing correlation heatmaps generated using data from male (left) and female (right) patients in the control group of GE 79-1.

to be related to ischemia [11]. Additionally, the slight differences in inflammation biomarkers between males and females seen in the heatmaps supports this explanation.

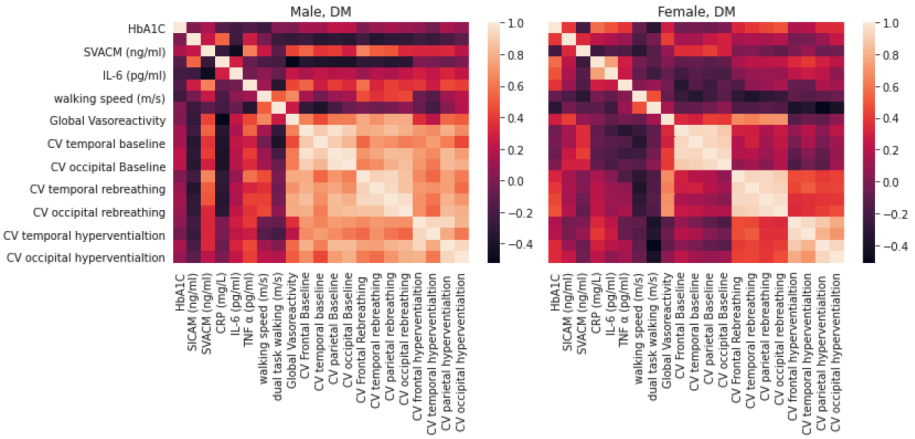

Figure 9: Comparing correlation heatmaps generated using data from male (left) and female (right) patients in the T2DM group of GE 79-1.

4.2 Bayesian Networks

Utilizing the Chow-Liu Bayesian network algorithm, it was found that severity of diabetes (measured by HbA1c %), SVCAM, and CRP were the cause of many other features and had significant connections throughout the network. In particular, severity was found to be the root cause of *all* the other features in every graph we generated. Additionally, walking speed was related to these three factors relatively strongly, and parietal baseline cerebral vasoreactivity (CV Par B) also seemed to have an impact among the features; however, measures of cerebral vasoreactivity were not included in GE 71 and 75 and thus limited further exploration of the feature's impact.

While the Chow-Liu algorithm is less accurate than other Bayesian network algorithms, it is the fastest, which was necessary to run the model in the coding environment (Google Colab). It is possible that other Bayesian network algorithms may produce more accurate or definite results. Such algorithms include the greedy, exact, and exact A* algorithms.

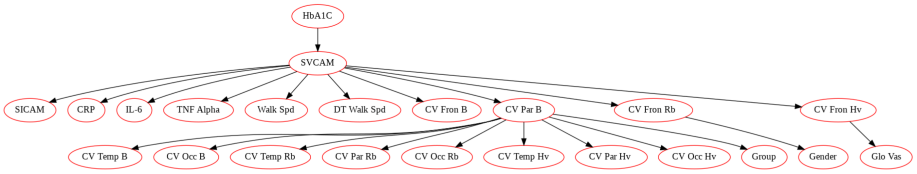

Figure 10: Bayesian network graph generated using the Chow-Liu algorithm on GE 79-1. Note the significance of HbA1c % and SVCAM.

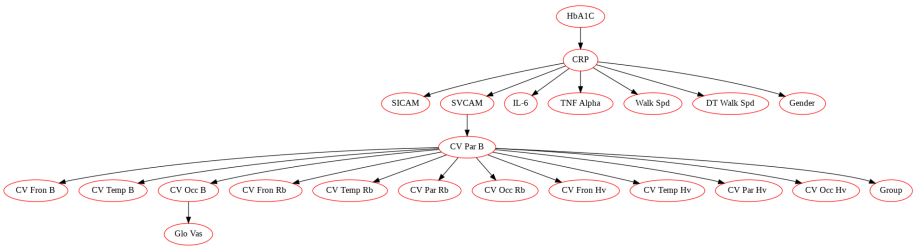

Figure 11: Bayesian network graph generated using the Chow-Liu algorithm on the T2DM Group of GE 79-1. Note the significance of HbA1c %, CRP, and SVCAM.

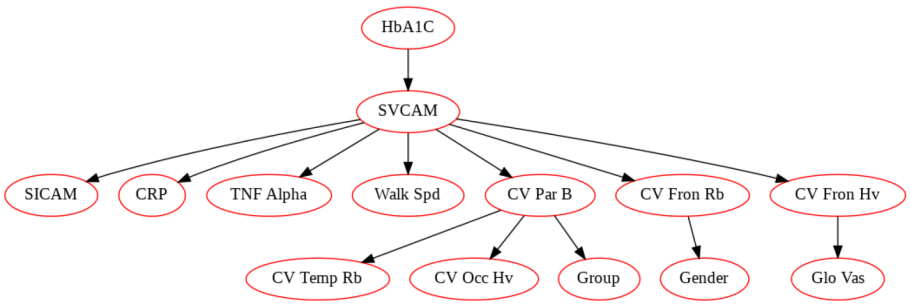

Figure 12: Bayesian network graph generated using the Chow-Liu algorithm on a subset of GE 79-1. Nodes for this graph were chosen based on features that seemed significant in Figure 10. Note that Figure 10 and Figure 12 have the same overall structure, with the difference being that Figure 12 has less nodes.

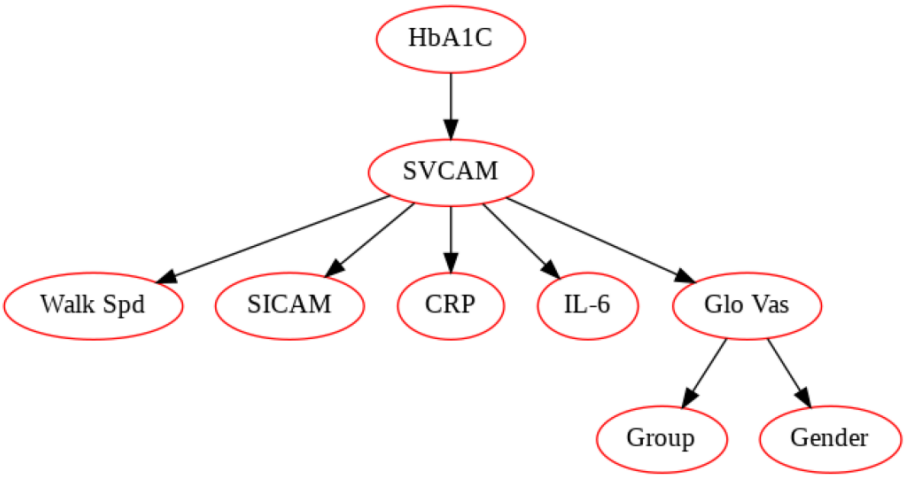

Figure 13: Bayesian network graph generated using the Chow-Liu algorithm on GE 75 and 79-1. Note the significance of HbA1c % and SVCAM.

We offer possible explanations for why severity, SVCAM, and CRP were

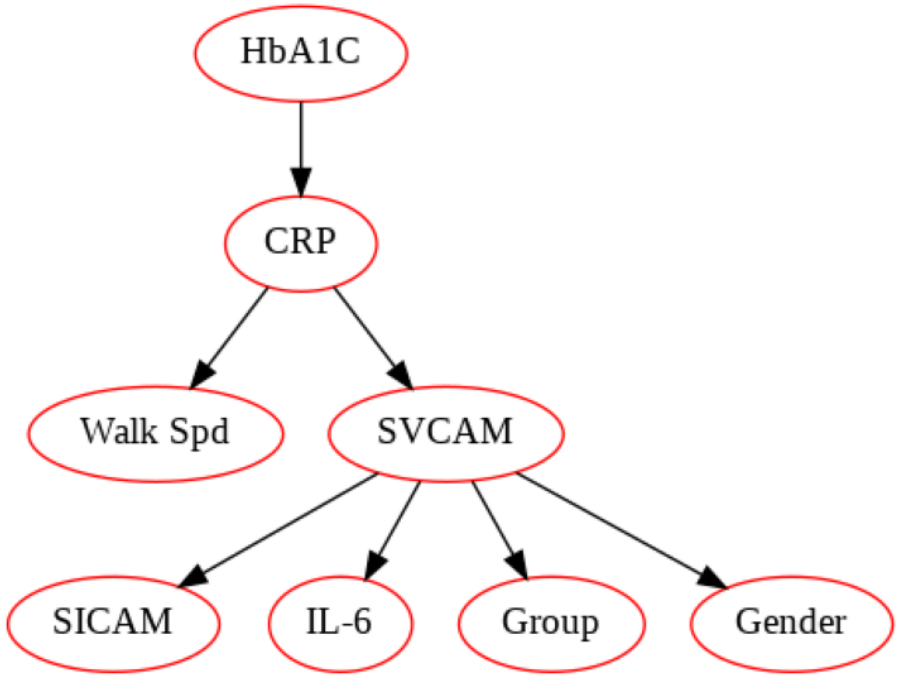

Figure 14: Bayesian network graph generated using the Chow-Liu algorithm on GE 71, 75, and 79-1. Note the significance of HbA1c %, CRP, SVCAM, and walking speed.

found to be the most significant causes in the Bayesian networks.

1. Severity is measured by HbA1c % and is directly related to the group of a patient (control or T2DM group). Since T2DM is a cause of CVD and the two diseases are strongly related to many of the features analyzed, it is probable that severity would have a large role.

2. SVCAM was the inflammation biomarker most related to the symptoms of CVD, having ties to hypertension, cerebral inflammation, cerebral vasoreactivity, and endothelial cell dysfunction. With these connections, SVCAM would likely be related to many features of CVD.

3. CRP is a protein used as a non-specific inflammation biomarker, and CRP tests are often conducted by medical practitioners to determine if inflammation is occurring as a result of disease. Due to its non-specific nature, CRP could be considered a more broad inflammation biomarker and thus would have many connections to different features.

4.3 Granger-Inspired Method

The results of the Granger-inspired method aligned with the results from the Bayesian networks. It was found that severity (HbA1c %), SVCAM, and walking speed all had a factor in causing IL-6, and that CRP, IL-6, and walking

speed all had a factor in causing severity. Each MSE ratio is significantly greater than 1. Overall, the features IL-6, CRP, severity, SVCAM, and walking speed were all connected to each other.

Table 5: The three groups of A, B, and C that had the highest MSE ratio for each model used are shown, as well as the MSE values for the AB Test group and the ABC test group.

Model	A, B, C	AB Test	ABC Test	AB/ABC MSE Ratio
Linear	A: IL-6 B: CRP C: HbA1c %	0.7248	0.3920	1.8490
Linear	A: IL-6 B: CRP C: SVCAM	0.7248	0.4756	1.5240
Linear	A: IL-6 B: CRP C: Walking Spd	0.7248	0.4780	1.5163
SVM	A: HbA1c % B: SVCAM C: CRP	2.4998	1.7742	1.4089
SVM	A: HbA1c % B: SVCAM C: IL-6	2.4998	1.7746	1.4086
SVM	A: HbA1c % B: SVCAM C: Walking Spd	2.4998	1.7752	1.4082

Table 5 shows the results of running both a linear regression model and an SVM model on different groups of A, B, and C. The first three groups of features with the highest MSE ratio are shown for each regression model, and the MSE values for the AB and ABC tests are shown as well.

While it is unclear why the linear regression and SVM model produced different results and ratios, we offer possible explanations for why IL-6 and walking speed were found to be significant using the Granger-inspired method.

1. IL-6 is an inflammation biomarker that can be used as a sign of chronic inflammation. Due to its non-specific nature, IL-6 may have connections for reasons similar to CRP.

2. Walking speed has been shown through multiple studies [5, 28] to be strongly related to lower cerebral vasoreactivity in patients with T2DM. Since walking speed is dependent on many factors in the brain, it is likely to be related to many features of CVD.

5 Conclusion

In this study, both machine learning methods of causal inference and traditional statistical tests were utilized by implementing a Bayesian network algorithm and a Granger causality-inspired approach as well as a linear regression method. These procedures were applied to find relationships among severity of T2DM, gender, inflammation biomarkers, and cerebral hypoperfusion. Overall, severity of T2DM, SVCAM, CRP, IL-6, and walking speed were the features found to have the strongest connections among all features examined. Additionally, there are significant differences between the correlation in cerebral vasoreactivities for males and females, particularly in the group with T2DM.

Using this information, testing for HbA1c %, SVCAM, CRP, and IL-6 levels in the blood could possibly aid medical practitioners in finding early signs of diseases related to CVD. Since tests already exist for HbA1c % and CRP, physicians may now be able to interpret the results of these tests in a new way. Also, since T2DM can cause CVD, and T2DM-induced CVD increases a patient's risk of other diseases, T2DM patients could be tested for these features (such as by measuring inflammation biomarker levels or by walking speeds) to find early signs of conditions such as cognitive decline, Alzheimer's, or stroke. This may allow physicians to find conditions before severe symptoms appear and will allow patients to receive treatment before diseases develop and worsen.

6 Future Work

6.1 Modifying the Study

For future work, a Bayesian network algorithm other than the Chow-Liu algorithm could be implemented to find if different or more detailed connections appear among the features. Applying a causal inference method other than a Bayesian network may also produce meaningful results. Also, if more data were to be collected regarding this research in the future, more significant results could be obtained. Finally, features of CVD other than those explored in this study could be considered, such as damage to endothelial cells, vessel wall thickening, and thinning of the blood-brain barrier [2].

6.2 Questions for Future Research

1. Why are severity of T2DM, SVCAM, CRP, and walking speed more significant than other features of CVD? Additionally, more studies could be conducted to determine how strong of a connection these four features have with one another.

2. Why are there significant differences among the correlations in cerebral vasoreactivities in the male and female brain? Research in an adjacent field, such as CVD-related stroke, could aid in exploring this question further.

3. How can connections or causal relationships found among these features be applied to the medical field and be used to better the lives of patients?

References

[1] M. A. B. Khan, M. J. Hashim, J. K. King, R. D. Govender, H. Mustafa, and J. Al Kaabi, "Epidemiology of type 2 diabetes – global burden of disease and forecasted trends," *Journal of Epidemiology and Global Health*, vol. 10, pp. 107–111, Mar. 2020.

[2] M. C. Zanon Zotin, L. Sveikata, A. Viswanathan, and P. Yilmaz, "Cerebral small vessel disease and vascular cognitive impairment: from diagnosis to management," *Current Opinion in Neurology*, vol. 34, pp. 246–257, Apr. 2021.

[3] M. S. Spychala, P. Honarpisheh, and L. D. McCullough, "Sex differences in neuroinflammation and neuroprotection in ischemic stroke," *Journal of Neuroscience Research*, vol. 95, pp. 462–471, Jan. 2017.

[4] M. Selim, R. Jones, P. Novak, P. Zhao, and V. Novak, "The effects of body mass index on cerebral blood flow velocity," *Clinical Autonomic Research*, vol. 18, pp. 331–338, Dec. 2008.

[5] A. J. Jor'dan, B. Manor, and V. Novak, "Slow gait speed – an indicator of lower cerebral vasoreactivity in type 2 diabetes mellitus," *Frontiers in Aging Neuroscience*, vol. 6, p. 135, 2014.

[6] C.-C. Chung, D. Pimentel, A. J. Jor'dan, Y. Hao, W. Milberg, and V. Novak, "Inflammation-associated declines in cerebral vasoreactivity and cognition in type 2 diabetes," *Neurology*, vol. 85, pp. 450–458, Aug. 2015.

[7] J. Duncombe, A. Kitamura, Y. Hase, M. Ihara, R. N. Kalaria, and K. Horsburgh, "Chronic cerebral hypoperfusion: a key mechanism leading to vascular cognitive impairment and dementia. Closing the translational gap between rodent models and human vascular cognitive impairment and dementia," *Clinical Science*, vol. 131, pp. 2451–2468, Oct. 2017.

[8] V. Novak and I. Hajjar, "The relationship between blood pressure and cognitive function," *Nature Reviews Cardiology*, vol. 7, pp. 686–698, Dec. 2010.

[9] V. Novak, F. Gomez, A. C. Dias, D. A. Pimentel, and F. J. Alfaro, "Diabetes-Related Cognitive Decline, a Global Health Issue, and New Treatments Approaches," *International Journal of Privacy and Health Information Management*, vol. 5, pp. 58–70, Dec. 2017.

[10] CDC, "All About Your A1C," Aug. 2018.

[11] A. E. Tchalla, G. A. Wellenius, T. G. Travison, M. Gagnon, I. Iloputaife, T. Dantoine, F. A. Sorond, and L. A. Lipsitz, "Circulating Vascular Cell Adhesion Molecule-1 Is Associated With Cerebral Blood Flow Dysregulation, Mobility Impairment, and Falls in Older Adults," *Hypertension*, vol. 66, pp. 340–346, Aug. 2015.

[12] B. Glowinska, M. Urban, J. Peczynska, and B. Florys, "Soluble adhesion molecules (sICAM-1, sVCAM-1) and selectins (sE selectin, sP selectin, sL selectin) levels in children and adolescents with obesity, hypertension, and

diabetes," *Metabolism: Clinical and Experimental*, vol. 54, pp. 1020–1026, Aug. 2005.

[13] A. Landry, P. Docherty, S. Ouellette, and L. J. Cartier, "Causes and outcomes of markedly elevated C-reactive protein levels," *Canadian Family Physician Medecin De Famille Canadien*, vol. 63, pp. e316–e323, June 2017.

[14] T. Tanaka, M. Narazaki, and T. Kishimoto, "IL-6 in Inflammation, Immunity, and Disease," *Cold Spring Harbor Perspectives in Biology*, vol. 6, pp. a016295–a016295, Oct. 2014.

[15] H. T. Idriss and J. H. Naismith, "TNF alpha and the TNF receptor superfamily: structure-function relationship(s)," *Microscopy Research and Technique*, vol. 50, pp. 184–195, Aug. 2000.

[16] Y. Chen, W. Duan, P. Sehrawat, V. Chauhan, F. J. Alfaro, A. Gavrieli, X. Qiao, V. Novak, and W. Dai, "Improved perfusion pattern score association with type 2 diabetes severity using machine learning pipeline: Pilot study: Perfusion Pattern Associates With T2DM," *Journal of Magnetic Resonance Imaging*, vol. 49, pp. 834–844, Mar. 2019.

[17] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Frontiers in Genetics*, vol. 9, p. 515, Nov. 2018.

[18] J. Suzuki, "A novel Chow–Liu algorithm and its application to gene differential analysis," *International Journal of Approximate Reasoning*, vol. 80, pp. 1–18, Jan. 2017.

[19] M. Horný and M.-Y. Lin, "Bayesian Networks," 2014.

[20] S. Kleinberg and G. Hripcsak, "A review of causal inference for biomedical informatics," *Journal of Biomedical Informatics*, vol. 44, pp. 1102–1112, Dec. 2011.

[21] V. Novak and L. Mendez, "Cerebral Vasoregulation in Diabetes." Type: dataset.

[22] V. Novak, D. Last, D. C. Alsop, A. M. Abduljalil, K. Hu, L. Lepicovsky, J. Cavallerano, and L. A. Lipsitz, "Cerebral blood flow velocity and periventricular white matter hyperintensities in type 2 diabetes," *Diabetes Care*, vol. 29, pp. 1529–1534, July 2006.

[23] Novak, Vera, Quispe, Rodrigo, and Saunders, Charles, "Cerebral perfusion and cognitive decline in type 2 diabetes." Type: dataset.

[24] V. Novak, P. Zhao, B. Manor, E. Sejdic, D. Alsop, A. Abduljalil, P. K. Roberson, M. Munshi, and P. Novak, "Adhesion molecules, altered vasoreactivity, and brain atrophy in type 2 diabetes," *Diabetes Care*, vol. 34, pp. 2438–2441, Nov. 2011.

[25] Novak, Vera and Quispe, Rodrigo, "Cerebromicrovascular Disease in Elderly with Diabetes." Type: dataset.

[26] V. Novak, P. Zhao, B. Manor, E. Sejdic, D. Alsop, A. Abduljalil, P. K. Roberson, M. Munshi, and P. Novak, "Adhesion molecules, altered vasoreactivity, and brain atrophy in type 2 diabetes," *Diabetes Care*, vol. 34, pp. 2438–2441, Nov. 2011.

[27] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, June 2000.

[28] C.-C. Chung, D. A. Pimentel Maldonado, A. J. Jor'dan, F. J. Alfaro, V.-A. Lioutas, M. Z. Núñez, and V. Novak, "Lower cerebral vasoreactivity as a predictor of gait speed decline in type 2 diabetes mellitus," *Journal of Neurology*, vol. 265, pp. 2267–2276, Oct. 2018.